Proposal

KAI CHANG, SAMI JAMAN, JUSTIN LEDERER, and OLLY STEWART, CMPU 250, Vassar College

1 INTRODUCTION

The movie business is a big business. According to Box Office Mojo, the US domestic box office gross for 2025 has already exceeded $986 million; each of the two preceding years, the total annual gross fell between $8.5 and $9 billion [9]. These numbers exclude international box office revenues, money spent to physically or digitally purchase or rent movies, and money spent to watch movies through subscription streaming services. In an environment where various new releases compete for the biggest slice of the box office pie, film studios and researchers are acutely interested in predicting which films will be financially successful and orienting release calendars accordingly. Research by Lash and Zhao [7] and Lee et al. [8], among others, has explored the use of machine learning techniques to predict movie success and profitability.

In the early 2000s, various research explored the effect of critical reviews on film success, and at least one study (Basuroy et al. [1]) found that "each of the first eight weeks, both positive and negative reviews are significantly correlated with box office revenue" [6] [1]. However, with the rise of the internet, consumers no longer turn to a Roger Ebert or Pauline Kael when deciding whether to see a film; instead, word-of-mouth (WOM) information on social media sites and crowd-sourced movie ratings like those from IMDb are among the first information presented to prospective filmgoers. Recent research has examined the effect of online WOM on movies and analyzed online review data, including using sentiment analysis to measure online response to a film [12] [13] [10].

A fundamental difference between crowdsourced and critical movie evaluation is the inherent manipulability of the former. Online communities, working in concert, can shape online movie scores by leaving masses of (usually negative) reviews to lower a movie's score, a phenomenon known as "review bombing." Some recent research has examined this phenomenon, including papers by Schuff et al. [11] and Cantone et al. [2]. Review bombing is most evident in the scores of films which are perceived as having made "woke" an existing (white and/or male) movie or franchise by rebooting it with women and/or people of color in the main roles (i.e. *Ghostbusters* (2016), *The Little Mermaid* (2023)) and films which address controversial topics like the Israel-Palestine conflict or the Armenian Genocide (i.e. *Israelism* (2023), *The Promise* (2016)). The prevalence of review bombing has led sites like IMDb to weight their ratings, leading to significant gaps between unweighted means and the publicly presented score (5.2 vs 7.2, 5.2 vs 6.8, 5.8 vs 7.7, and 6.0 vs 6.1 for each of the previously mentioned movies, respectively. The smaller gap for The Promise is due to the occurrence of simultaneous positive review bombing, a phenomenon also visible for some of the other films.)

When online reviews can shape film success and by extension future studio decisions, review bombing is a potentially potent tool to punish filmmakers and studios who make brave or unpopular decisions. Our project explores review bombing, analyzing which kinds of films are more likely to review bombed, how that relates to their success, how platforms like IMDb use weighting to minimize the effects of review bombing, and how ratings for review bombed films differ in different countries.

Some important terms it is necessary to define for our project are as follows: polarization, polarization score, international polarization, weighted rating, unweighted rating, 2-9 rating. Some of these are fairly simple; the *weighted rating* is a film's basic overall rating provided by IMDb, while the *unweighted rating* is the unweighted mean rating provided on a film's IMDb page. 2-9 rating refers to the unweighted mean rating of a film when extreme scores (1s and 10s) are removed and only scores 2-9 are considered. *Polarization* refers to the amount of disagreement in a film's ratings; a film is more polarized if it has a higher *polarization score*. We calculate a film's polarization score using a version of the polarization formula created by Esteban and Ray [3]; the maximum score a film can receive is 2.25, meaning that exactly half the ratings are 10 and the other half are 1, and the lowest is 0, if all the ratings are identical (Hickey and Mehta [4] also use a version of this formula, but a slightly different one). *International polarization* of a film between two countries refers to the difference in unweighted mean rating for the two countries. These are some of the basic metrics we will use to identify films that have been review bombed and analyze the extent and nature of the review bombing.

1.1 Research Questions

In this project, we aim to explore the following research questions:

(RQ1) Movies belonging to which genres are more likely to have polarized rating distributions?
(RQ2) Which movies' average ratings differ the most across different countries?
(RQ3) What types of cast demographics are correlated with movies with more polarized rating distributions?

2 DATA DESCRIPTION

The data is from IMDb and was collected by scraping their website and querying their GraphQL API. Each observation refers to a movie. There are observations in the data for each movie that has over 50,000 IMDb user ratings (as of 2/24/25), for a total of 4,329 observations. Each observation contains two nominal variables (`id`, `title`), a comma-delimited set of categorical variables for the movie's genres (`genres`), a numeric variable for the movie's release year (`release_year`), and various rating-related variables, described in Table 1.

Variable	Description
`weighted_rating`	A continuous numeric variable for the weighted average of user votes calculated using a proprietary algorithm. IMDb uses an alternative algorithm when "unusual voting activity is detected."[5]
`rating_{1...10}`	A discrete numeric variable for the total number of votes for each rating.
`country_{0...4}`	A nominal variable for each of the five countries with the most votes cast.
`country_{0...4}_rating_{1...10)`	A discrete numeric variable for the number of votes for each rating for each of the five countries with the most votes cast.

Table 1. Descriptions of rating-related variables

2.1 Description of what you expect to find

In our exploratory data analysis, we will measure the polarization of rating distributions using two measures, first applied to movie reviews by Hickey and Mehta at FiveThirtyEight in 2017 [4]: Esteban and Ray's polarization score [3] and the proportion of "polarized" votes, ones that are either "1" or "10." We will calculate the difference in unweighted mean rating for the two countries with the greatest difference out of the five countries with the most votes. We will measure the correlation between cast demographics and more polarized rating distributions using a linear regression.

We expect to find that movies belonging to genres that may contain politically controversial subject material ("History," "Biography," and "Documentary," for example) will be more likely to have polarized rating distributions and to have average ratings that differ the most across different countries. Examples of individual movies that support this hypothesis mentioned in the introduction are *The Promise* (2016) and *Israelism* (2023) (though *Israelism* will not be included in our initial analysis because it has fewer than 50,000 user votes). We also expect to find that movies with higher proportions of female and non-white stars will be more likely to have polarized rating distributions. Examples of individual movies that support this hypothesis mentioned in the introduction are *Ghostbusters* (2016) and *The Little Mermaid* (2023).

3 WORK AGREEMENT

Our group has agreed to meet weekly in order to discuss and check in on our individual project roles. At the meeting, each group member will offer a brief summary of what they have completed in the past week. They can also ask for help in specific aspects of their task, if they are unable to complete their portion on their own. The meetings will take place in the Deece at 10:30-11:45 AM. and additional meetings may be scheduled at the discretion of group members should we deem it necessary. The portion of the work that each group member is responsible for is expected to be completed independently by the following meeting time on Tuesday. In the case that any individual needs help earlier and needs to reach the group, we have created a shared group chat to be used in order to get into contact with fellow group members. This will allow us to maintain continuous and smooth communication so that we can adjust our plans and help each other stay productive and on schedule. Our goal with setting Tuesday deadlines for ourselves is to ensure that we have adequate time to edit our work before the class deadline rather than scrambling to complete parts of the project last-minute. In addition, all group members are expected to produce consistent, high-quality work and maintain clear and open communication, so we can adjust work assignments and expectations when necessary.

REFERENCES

[1] Suman Basuroy, Subimal Chatterjee, and S. Abraham Ravid. 2003. How Critical are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing* 67, 4 (2003), 103–107.

[2] Giulio Giacomo Cantone, Venera Tomaselli, and Valeria Mazzeo. 2024. Review bombing: ideology-driven polarisation in online ratings: The case study of The Last of Us (part II). *Quality Quantity* (2024).

[3] Joan-Maria Esteban and Debraj Ray. 1994. On the Measurement of Polarization. *Econometrica* 62, 4 (1994), 819–851.

[4] Walt Hickey and Dhrumil Mehta. 2017. Al Gore's New Movie Exposes The Big Flaw In Online Movie Ratings. https://fivethirtyeight.com/features/al-gores-new-movie-exposes-the-big-flaw-in-online-movie-ratings.

[5] IMDb. [n.d.]. Weighted Average Ratings. https://help.imdb.com/article/imdb/track-movies-tv/weighted-average-ratings/GWT2DSBYVT2F25SK#. Accessed: 2024-02-25.

[6] Timothy King. 2007. Does film criticism affect box office earnings? Evidence from movies released in the U.S. in 2003. *Journal of Cultural Economics* 31 (2007), 171–186.

[7] Michael T. Lash and Kang Zhao. 2016. Early Predictions of Movie Success: The Who, What, and When of Profitability. *Journal of Management Information Systems* 33, 3 (2016), 874–903.

[8] Kyuhan Lee, Jinsoo Park, Iljoo Kim, and Youngseok Choi. 2018. Predicting movie success with machine learning techniques: ways to improve accuracy. *Information Systems Frontiers* 20 (2018), 577–588.

[9] Box Office Mojo. [n.d.]. Domestic Yearly Box Office. https://www.boxofficemojo.com/year. Accessed: 2024-02-25.

[10] Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten de Rijke. 2012. Predicting IMDB Movie Ratings Using Social Media. In *Advances in Information Retrieval*. 503–507.

[11] David Schuff, Susan Mudambi, and Mei-Xian Wang. 2024. Understanding the Review Bombing Phenomenon in Movies and Television. In *57th Hawaii International Conference on System Sciences*.

[12] Harsh Sharma, Satyajit Pangaonkar, Reena Gunjan, and Prakash Rokade. 2003. Sentimental Analysis of Movie Reviews Using Machine Learning. In *ITM Web of Conferences*, Vol. 53.

[13] Yeujun Yoon, Charin Polpanumas, and Young Joon Park. 2017. The Impact of Word of Mouth via Twitter On Moviegoers' Decisions and Film Revenues. *Journal of Advertising Research* 57, 2 (2017), 144–158.